Individual Assignment 2: Visualizing Time

Chris Anderson
PREDICT 455 Section 56

Summary and Problem Definition

Much has been written about Chicago Cubs pitcher Jake Arrieta's rise to the elite levels of baseball performance. His performance improved significantly after he was traded from the Baltimore Orioles to the Chicago Cubs, culminating in a historic run of success during the 2015 season, and it has been reported that the improvement stems from a change in his release point (Arthur, 2015). Some have pointed specifically to his vertical release point as a key to his success (Altman, 2016). Using a line graph to visualize his release point over time illustrates how much it has actually changed and if it has actually had an impact.

Methods

PITCHf/x is a motion tracking system that captures data on the release point, speed, and trajectory of every pitch that occurs in every game throughout the league. Using MLB.com's Statcast search tool PITCHf/x data for every pitch thrown by Jake Arrieta over the course of his career (2010 to 2016) was downloaded in the form of a.csv file. R was used to perform minimal data preparation and calculate mean vertical release point and horizontal release point per year. These mean values were joined with another dataset obtained from the Statcast search tool containing Arrieta's results (e.g. Batting Average Against) for each year. Ultimately, the results were loaded into Tableau to produce the visualization.

Programming Overview

Very little programming was actually needed to prepare the data for use in Tableau. Two files were read into R: the first containing the PITCHf/x data (including release point measurements) for all pitches thrown in Arrieta's career, and the second containing his performance measurements or "results" for each year. The *aggregate* function was used on the data from the first file to calculate the mean vertical

release point and mean horizontal release point per year. These values were merged with the second file containing the results for each year and the resulting table was written to a .csv file, which served as the data source in Tableau. Three vertically-stacked line graphs were produced, each with its own y-axis but sharing a common x-axis. This allows for easy comparison but avoid the confusion of multiple y-axes in the same space. An important note is the y-axes do not have a zero baseline. This is because the numbers are very small and slight, but extremely important in the context of baseball, changes were not evident in the graph without adjusting the baseline.

Results

The colors in the graphic make it apparently immediately that Arrieta changed something soon after his arrival in Chicago. The changes in horizontal release point track closely to the changes in batting average against throughout his career, while his vertical release point has risen steadily. Looking at the 2016 season, there is a slight increase in Batting Average Against, accompanied by increases in **both** horizontal and vertical release point. While the horizontal release increased only slightly, the vertical release point ascended to values it had never before reached. It may be far to wonder if this fluctuation will continue being a non-factor as it soars past its previous heights.

References

Altman, E. (2016, June 07). I Think I May Have Found Something 'Wrong' with Jake Arrieta - Cubs Insider. Retrieved October 16, 2016, from http://www.cubsinsider.com/something-wrong-jake-arrieta/

Arthur, R. (2015, September 14). Jake Arrieta Is Not Your Normal Ace. Retrieved October 16, 2016, from http://fivethirtyeight.com/features/jake-arrieta-is-not-your-normal-ace/

Statcast Search. (2016). Retrieved July 10, 2016, from https://baseballsavant.mlb.com/statcast_search